DOI:
10.3115/1219840.1219869

Document Version
Peer reviewed version

Citation for published version (APA):
Ginzburg, J., & Fernandez, R. (2005). Scaling up from Dialogue to Multilogue: some principles and benchmarks. In ACL '05 Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics. (pp. 231-238). Stroudsburg, PA, USA: Association for Computational Linguistics. 10.3115/1219840.1219869

Scaling up from Dialogue to Multilogue: some principles and benchmarks

Jonathan Ginzburg and Raquel Fernández
Dept of Computer Science
King's College, London
The Strand, London WC2R 2LS
UK
{ginzburg,raquel}@dcs.kcl.ac.uk

Abstract

The paper considers how to scale up dialogue protocols to multilogue, settings with multiple conversationalists. We extract two benchmarks to evaluate scaled up protocols based on the long distance resolution possibilities of non-sentential utterances in dialogue and multilogue in the British National Corpus. In light of these benchmarks, we then consider three possible transformations to dialogue protocols, formulated within an issue-based approach to dialogue management. We show that one such transformation yields protocols for querying and assertion that fulfill these benchmarks.

1 Introduction

The development of dialogue systems in which a human agent interacts using natural language with a computational system is by now a flourishing domain (see e.g. (NLE, 2003)), buttressed by an increasing theoretical and experimental literature on the properties of dialogue (see e.g. recent work in the SEMDIAL and SIGDIAL conferences). In contrast, the development of *multilogue* systems, in which conversation with 3 or more participants ensue—is still in its early stages, as is the theoretical and experimental study of multilogue. *The* fundamental issue in tackling multilogue is: how can mechanisms motivated for dialogue (e.g. information states, protocols, update rules etc) be scaled up to multilogue?

In this paper we extract from a conversational corpus, the British National Corpus (BNC), several benchmarks that characterize dialogue and multilogue interaction. These are based on the resolution possibilities of non-sentential utterances (NSUs). We then use these benchmarks to evaluate certain general transformations whose application to a dialogue interaction system yield a system appropriate for multilogue.

There are of course various plausible views of the relation between dialogue and multilogue. One possible approach to take is to view multilogue as a sequence of dialogues. Something like this approach seems to be adopted in the literature on communication between autonomous software agents. However, even though many situations considered in multiagent systems do involve more than two agents, most interaction protocols are designed only for two participants at a time. This is the case of the protocol specifications provided by FIPA (Foundation for Intelligent Physical Agents) for agent communication language messages (FIPA, 2003). The FIPA interaction protocols (IP) are most typically designed for two participants, an initiator and a responder . Some IPs permit the broadcasting of a message to a group of addressees, and the reception of multiple responses by the original initiator (see most particularly the Contract Net IP). However, even though more than two agents participate in the communicative process, as (Dignum and Vreeswijk, 2003) point out, such conversations can not be considered multilogue, but rather a number of parallel dialogues.

The Mission Rehearsal Exercise (MRE) Project (Traum and Rickel, 2002), one of the largest multilogue systems developed hitherto, is a virtual reality environment where multiple partners (including humans and other autonomous agents) engage in multi-conversation situations. The MRE is underpinned by an approach to the modelling of interaction in terms of obligations that different utterance types bring about originally proposed for dialogue (see e.g. (Matheson et al. , 2000)). In particular, this includes a model of the grounding process (Clark, 1996) that involves recognition and construction of common ground units (CGUs) (see (Traum, 2003)). Modelling of obligations and grounding becomes more complex when considering multilogue situations. The model of grounding implemented in the MRE project can only be used in cases where there is a single initiator and responder. It is not clear what the model should be for multiple addressees: should the contents be considered grounded when any of the addressees has acknowledged them? Should evidence of understanding be required from every addressee?

Since their resolution is almost wholly reliant on context, non sentential utterances provide a large testbed concerning the structure of both dialogue and multilogue. In section 2 we present data from the British National Corpus (BNC) concerning the resolution of NSUs in dialogue and multilogue. The main focus of this data is with the distance between antecedent and fragment. We use this to extract certain benchmarks concerning multilogue interaction. Thus, acknowledgement and acceptance markers (e.g. 'mmh', 'yeah') are resolved with reference to an utterance (assertion) which they ground (accept). The data we provide shows that acknowledgements in multilogue, as in dialogue, are adjacent to their antecedent. This provides evidence that, in general, a single addressee serves to signal grounding. In contrast, BNC data indicates the prevalence in multilogue of short answers that are resolved using material from an antecedent question located several turns back, whereas in dialogue short answers are generally adjacent to their antecedent. This provides evidence against reducing querying interaction in multilogue to a sequence of dialogues. We show that long distance short answers are a stable phenomenon for multilogue involving both small ($\leq$5 persons) and large ($>$ 5 persons) groups, despite the apparently declining interactivity with increasing group size flagged in experimental work (see (Fay et al., 2000)).

In section 3 we sketch the basic principles of issue based dialogue management which we use as a basis for our subsequent investigations of multilogue interaction. This will include information states and formulation of protocols for querying and assertion in dialogue. In section 4 we consider three possible transformations on dialogue protocols into multilogue protocols. These transformations are entirely general in nature and could be applied to protocols stated in whatever specification language. We evaluate the protocols that are generated by these transformations with reference to the benchmarks extracted in section 2. In particular, we show that one such transformation, dubbed Add Side Participants(ASP), yields protocols for querying and assertion that fulfill these benchmarks. Finally, section 5 provides some conclusions and pointers to future work.

2 Long Distance Resolution of NSUs in Dialogue and Multilogue: some benchmarks

The work we present in this paper is based on empirical evidence provided by corpus data extracted from the British National Corpus (BNC).

2.1 The Corpus

Our current corpus is a sub-portion of the BNC conversational transcripts consisting of 14,315 sentences. The corpus was created by randomly excerpting a 200-speaker-turn section from 54 BNC files. Of these files, 29 are transcripts of conversations between two dialogue participants, and 25 files are multilogue transcripts.

A total of 1285 NSUs were found in our sub-corpus. Table 1 shows the raw counts of NSUs found in the dialogue and multilogue transcripts, respectively.

	NSUs	BNC files
Dialogue	709	29
Multilogue	576	25
Total	1285	54

Table 1: Total of NSUs in Dialogue and Multilogue

All NSUs encountered within the corpus were classified according to the NSU typology presented in (Fernández and Ginzburg, 2002). Additionally, the distance from their antecedent was measured.[1] Table 2 shows the distribution of NSU categories and their antecedent separation distance. The classes of NSU which feature in our discussion below are boldfaced.

The BNC annotation includes tagging of units approximating to sentences, as identified by the CLAWS segmentation scheme (Garside, 1987). Each sentence unit is assigned an identifier number. By default it is assumed that sentences are non-overlapping and that their numeration indicates temporal sequence. When this is not the case because speakers overlap, the tagging scheme encodes synchronous speech by means of an alignment map used to synchronize points within the transcription. However, even though information about simultaneous speech is available, overlapping sentences are annotated with different sentence numbers.

In order to be able to measure the distance between the NSUs encountered and their antecedents, all instances were tagged with the sentence number of their antecedent utterance. The distance we report is therefore measured in terms of sentence numbers. It should however be noted that taking into account synchronous speech would not change the data reported in Table 2 in any significant

[1] This classification was done by one expert annotator. To assess its reliability a pilot study of the taxonomy was performed using two additional non-expert coders. These annotated 50 randomly selected NSUs (containing a minimum of 2 instances of each NSU class, as labelled by the expert annotator.). The agreement achieved by the three coders is reasonably good, yielding a kappa score κ = 0.76. We also assessed the accuracy of the coders' choices in choosing the antecedent utterance using the expert annotator's annotation as a gold standard. Given this, one coder's accuracy was 92%, whereas the other coder's was 96%.

			Distance						
NSU Class	**Example**	**Total**	1	2	3	4	5	6	>6
Acknowledgment	*Mm mm.*	595	578	15	2				
Short Answer	*Ballet shoes.*	188	104	21	17	5	5	8	28
Affirmative Answer	*Yes.*	109	104	4			1		
Clarification Ellipsis	*John?*	92	76	13	2	1			
Repeated Ack.	*His boss, right.*	86	81	2	3				
Rejection	*No.*	50	49	1					
Factual Modifier	*Brilliant!*	27	23	2	1	1			
Repeated Aff. Ans.	*Very far, yes.*	26	25	1					
Helpful Rejection	*No, my aunt.*	24	18	5		1			
Check Question	*Okay?*	22	15	7					
Filler	*... a cough.*	18	16	1		1			
Bare Mod. Phrase	*On the desk.*	16	11	4			1		
Sluice	*When?*	11	10	1					
Prop. Modifier	*Probably.*	11	10	1					
Conjunction Phrase	*Or a mirror.*	10	5	4	1				
	Total	1285	1125	82	26	9	7	8	28
	Percentage	100	87.6	6.3	2	0.6	0.5	0.6	2.1

Table 2: NSUs sorted by Class and Distance

way, as manual examination of all NSUs at more than distance 3 reveals that the transcription portion between antecedent and NSU does not contain any completely synchronous sentences in such cases.

In the examples throughout the paper we shall use italics to indicate speech overlap. When italics are not used, utterances take place sequentially.

2.2 NSU-Antecedent Separation Distance

The last row in Table 2 shows the distribution of NSU-antecedent separation distances as percentages of the total of NSUs found. This allows us to see that about 87% of NSUs have a distance of 1 sentence (i.e. the antecedent was the immediately preceding sentence), and that the vast majority (about 96%) have a distance of 3 sentences or less.

Although the proportion of NSUs found in dialogue and multilogue is roughly the same (see Table 1 above), when taking into account the distance of NSUs from their antecedent, the proportion of long distance NSUs in multilogue increases radically: the longer the distance, the higher the proportion of NSUs that were found in multilogue. In fact, as Table 3 shows, NSUs that have a distance of 7 sentences or more appear exclusively in multilogue transcripts. These differences are significant (χ^2 = 62.24, $p \leq 0.001$).

Adjacency of grounding and affirmation utterances The data in table 2 highlights a fundamental characteristic of the remaining majoritarian classes of NSUs, Ack(nowledgements), Affirmative Answer, CE (clarification ellipsis), Repeated Ack(nowledgements), and Rejection. These are used either in grounding interaction, or to affirm/reject propositions.[2] The overwhelming adjacency to their antecedent underlines the locality of these interactions.

Long distance potential for short answers One striking result exhibited in Table 2 is the uneven distribution of long distance NSUs across categories. With a few exceptions, NSUs that have a distance of 3 sentences or more are exclusively short answers. Not only is the long distance phenomenon almost exclusively restricted to short answers, but the frequency of long distance short answers stands in strong contrast to the other NSUs classes; indeed, over 44% of short answers have more than distance 1, and over 24% have distance 4 or more, like the last answer in the following example:

(1) Allan: How much do you think?
Cynthia: Three hundred pounds.
Sue: More.
Cynthia: A thousand pounds.
Allan: More.
Unknown: <unclear>
Allan: Eleven hundred quid apparently.
[BNC, G4X]

Long distance short answers primarily a multilogue effect Table 4 shows the total number of short answers found in dialogue and multilogue respectively, and the proportions sorted by distance over those totals:

From this it emerges that short answers are more common in multilogue than in dialogue—134(71%) v.

[2] Acknowledgements and acceptances are, in principle, distinct acts: the former involves indication that an utterance has been understood, whereas the latter that an assertion is accepted. In practice, though, acknowledgements in the form of NSUs commonly simultaneously signal acceptances. Given this, corpus studies of NSUs (e.g. (Fernández and Ginzburg, 2002)) often conflate the two.

Distance	1	2	3	4	5	6	>6
Dialogue	658 (59%)	37 (45%)	11 (45%)	1 (12%)	1 (14%)	1 (13%)	0 (0%)
Multilogue	467 (41%)	45 (55%)	15 (55%)	8 (88%)	6 (86%)	7 (87%)	28 (100%)

Table 3: NSUs in dialogue and multilogue sorted by distance

Short Answers	Total #	1	2	3	> 3
Dialogue	54	82	9	9	0
Multilogue	134	44	11	8	37

Table 4: % over the totals found in dialogue and multilogue

54(29%). Also, the distance pattern exhibited by these two groups is strikingly different: Only 18% of short answers found in dialogue have a distance of more than 1 sentence, with all of them having a distance of at most 3, like the short answer in (2).

(2) Malcolm: [...] cos what's three hundred and sixty divided by seven?
Anon 1: I don't know.
Malcolm: Yes I don't know either!
Anon 1: Fifty four point fifty one point four. [BNC, KND]

This dialogue/multilogue asymmetry argues against reductive views of multilogue as sequential dialogue.

Long Distance short answers and group size As Table 4 shows, all short answers at more than distance 3 appear in multilogues. Following (Fay et al., 2000), we distinguish between small groups (those with 3 to 5 participants) and large groups (those with more than 5 participants). The size of the group is determined by the amount of participants that are active when a particular short answer is uttered. We consider active participants those that have made a contribution within a window of 30 turns back from the turn where the short answer was uttered.

Table 5 shows the distribution of long distance short answers (distance > 3) in small and large groups respectively. This indicates that long distance short answers are significantly more frequent in large groups ($\chi^2 = 22.17$, $p \leq 0.001$), though still reasonably common in small groups. A pragmatic account correlating group size and frequency of long distance short answers is offered in the final paragraph of section 3.

Group Size	d > 3	d ≤ 3	Total
≤ 5	20 (21.5%)	73 (78.5%)	93
> 5	26 (63%)	15 (37%)	41

Table 5: Long distance short answers in small and large groups

Large group multilogues in the corpus are all transcripts of tutorials, training sessions or seminars, which exhibit a rather particular structure. The general pattern involves a question being asked by the tutor or session leader, the other participants then taking turns to answer that question. The tutor or leader acts as turn manager. She assigns the turn explicitly usually by addressing the participants by their name without need to repeat the question under discussion. An example is shown in (3):

(3) Anon1: How important is those three components and what value would you put on them [...]
Anon3: Tone forty five. Body language thirty .
Anon1: Thank you.
Anon4: Oh.
Anon1: Melanie.
Anon5: twenty five.
Anon1: Yes.
Anon5: Tone of voice twenty five. [BNC, JYM]

Small group multilogues on the other hand have a more unconstrained structure: after a question is asked, the participants tend to answer freely. Answers by different participants can follow one after the other without explicit acknowledgements nor turn management, like in (4):.

(4) Anon 1: How about finance then? <pause>
Unknown 1: Corruption
Unknown 2: Risk <pause dur=30>
Unknown 3: Wage claims <pause dur=18>

2.3 Two Benchmarks of multilogue

The data we have seen above leads in particular to the following two benchmarks protocols for querying, assertion, and grounding interaction in multilogue:

(5) a. **Multilogue Long Distance short answers (MLDSA)**: querying protocols for multilogue must license short answers an unbounded number of turns from the original query.

b. **Multilogue adjacency of grounding/acceptance (MAG)**: assertion and grounding protocols for multilogue should license grounding/clarification/acceptance moves only adjacently to their antecedent utterance.

MLDSA and MAG have a somewhat different status: whereas MLDSA is a direct generalization from the data, MAG is a negative constraint, posited given the paucity of positive instances. As such MAG is more open to doubt and we shall treat it as such in the sequel.

3 Issue based Dialogue Management: basic principles

In this section we outline some of the basic principles of Issue-based Dialogue Management, which we use as a basis for our subsequent investigations of multilogue interaction.

Information States We assume information states of the kind developed in the KoS framework (e.g. (Ginzburg, 1996, forthcoming), (Larsson, 2002)) and implemented in systems such as GODIS, IBIS, and CLARIE (see e.g. (Larsson, 2002; Purver, 2004)). On this view each dialogue participant's view of the common ground, their Dialogue Gameboard (DGB), is structured by a number of attributes including the following three: FACTS: a set of facts representing the shared assumptions of the CPs, LatestMove: the most recent grounded move, and QUD ('questions under discussion'): a partially ordered set—often taken to be structured as a stack—consisting of the currently discussable questions.

Querying and Assertion Both querying and assertion involve a question becoming maximal in the querier/asserter's QUD:[3] the posed question q for a query where q is posed, the polar question $p?$ for an assertion where p is asserted. Roughly, the responder can subsequently either choose to start a discussion (of q or $p?$) or, in the case of assertion, to update her FACTS structure with p. A dialogue participant can downdate $q/p?$ from QUD when, as far as her (not necessarily public) goals dictate, sufficient information has been accumulated in FACTS. The querying/assertion protocols (in their most basic form) are summarized as follows:

(6)

querying	assertion
LatestMove = Ask(A,q)	LatestMove = Assert(A,p)
A: push q onto QUD; release turn;	A: push p? onto QUD; release turn
B: push q onto QUD; take turn; make max-qud–specific; utterance[4] take turn.	B: push p? onto QUD; take turn; Option 1: Discuss p? Option 2: Accept p
	LatestMove = Accept(B,p)
	B: increment FACTS with p; pop p? from QUD;
	A: increment FACTS with p; pop p? from QUD;

Following (Larsson, 2002; Cooper, 2004), one can decompose interaction protocols into *conversational update rules*—functions from DGBs into DGBs using Type Theory with Records (TTR). This allows simple interfacing with the grammar, a Constraint-based Grammar closely modelled on HPSG but formulated in TTR (see (Ginzburg, forthcoming)).

Grounding Interaction Grounding an utterance $u : T$ ('the sign associated with u is of type T') is modelled as involving the following interaction. (a) Addressee B tries to anchor the contextual parameters of T. If successful, B acknowledges u (directly, gesturally or implicitly) and responds to the content of u. (b) If unsuccessful, B poses a Clarification Request (CR), that arises via *utterance coercion* (see (Ginzburg and Cooper, 2001)). For reasons of space we do not formulate an explicit protocol here—the structure of such a protocol resembles the assertion protocol. Our subsequent discussion of assertion can be modified *mutatis mutandis* to grounding.

NSU Resolution We assume the account of NSU resolution developed in (Ginzburg and Sag, 2000). The essential idea they develop is that NSUs get their main predicates from context, specifically via unification with the question that is currently *under discussion*, an entity dubbed the *maximal question under discussion* (MAX-QUD). NSU resolution is, consequently, tied to conversational topic, viz. the MAX-QUD.[5]

Distance effects in dialogue short answers If one assumes QUD to be a stack, this affords the potential for non adjacent short answers in dialogue. These, as discussed in section 2, are relatively infrequent. Two commonly observed *dialogue* conditions will jointly enforce adjacency between short answers and their interrogative antecedents: (a) Questions have a simple, one phrase answer. (b) Questions can be answered immediately, without preparatory or subsequent discussion. For multilogue (or at least certain genres thereof), both these conditions are less likely to be maintained: different CPs can supply different answers, even assuming that relative to each CP there is a simple, one phrase answer. The more CPs there are in a conversation, the smaller their common ground and the more likely the need for clarificatory interaction. A pragmatic account of this type of the frequency of adjacency in dialogue short answers seems clearly preferable to any actual mechanism that would *rule out* long distance short answers. These can be perfectly felicitous—see e.g. example (1) above which

[3] In other words, *pushed onto the stack*, if one assumes QUD is a stack.

[4] An utterance whose content is either a proposition p About max-qud or a question q_1 on which max-qud Depends. For the latter see footnote 7. If one assumes QUD to be a stack, then 'max-qud–specific' will in this case reduce to 'q–specific'. But the more general formulation will be important below.

[5] The resolution of NSUs, on the approach of (Ginzburg and Sag, 2000), involves one other parameter, an antecedent subutterance they dub the *salient-utterance* (SAL-UTT). This plays a role similar to the role played by the *parallel element* in higher order unification–based approaches to ellipsis resolution (see e.g. (Pulman, 1997). For current purposes, we limit attention to the MAX-QUD as the nucleus of NSU resolution.

would work fine if the turn uttered by Sue had been uttered by Allan instead. Moreover such a pragmatic account leads to the expectation that the frequency of long distance antecedents is correlated with group size, as indeed indicated by the data in table 5.

4 Scaling up Protocols

(Goffman, 1981) introduced the distinction between *ratified participants* and *overhearers* in a conversation. Within the former are located the speaker and participants whom she takes into account in her utterance design—the intended addressee(s) of a given utterance, as well as *side participants*. In this section we consider three possible principles of protocol extension, each of which can be viewed as adding roles for participants from one of Goffman's categories. We evaluate the protocol that results from the application of each such principle relative to the benchmarks we introduced in section 2.3. Seen in this light, the final principle we consider, Add Side Participants (ASP), arguably, yields the best results. Nonetheless, these three principles would appear to be complementary—the most general protocol for multilogue will involve, minimally, application of all three.[6] We state the principles informally and framework independently as transformations on operational construals of the protocols. In a more extended presentation we will formulate these as functions on TTR conversational update rules.

The simplest principle is Add Overhearers (AOV). This involves adding participants who merely observe the interaction. They keep track of facts concerning a particular interaction, but their context is not facilitated for them to participate:

(7) Given a dialogue protocol π, add roles $C_1,\ldots,C_n$ where each C_i is a silent participant: given an utterance u_0 classified as being of type T_0, C_i updates C_i.DGB.FACTS with the proposition $u_0 : T_0$.

Applying AOV yields essentially multilogues which are sequences of dialogues. A special case of this are moderated multilogues, where all dialogues involve a designated individual (who is also responsible for turn assignment.). Restricting scaling up to applications of AOV is not sufficient since *inter alia* this will not fulfill the MLDSA benchmark.

A far stronger principle is Duplicate Responders (DR):

(8) Given a dialogue protocol π, add roles $C_1,\ldots,C_n$ which duplicate the responder role.

[6] We thank an anonymous reviewer for ACL for convincing us of this point.

Applying DR to the querying protocol yields the following protocol:

(9) Querying with multiple responders

1. LatestMove = Ask(A,q)
2. A: push q onto QUD; release turn
3. Resp_1: push q onto QUD; take turn; make max-qud–specific utterance; release turn
4. Resp_2: push q onto QUD; take turn; make max-qud–specific utterance; release turn
5. . . .
6. Resp_n: push q onto QUD; take turn; make max-qud–specific utterance; release turn

This yields interactions such as (4) above. The querying protocol in (9) licenses long distance short answers, so satisfies the MLDSA benchmark. On the other hand, the contextual updates it enforces will not enable it to deal with the following (constructed) variant on (4), in other words does not afford responders to comment on previous responders, as opposed to the original querier:

(10) A: Who should we invite for the conference?
B: Svetlanov.
C: No (=Not Svetlanov), Zhdanov
D: No (= Not Zhdanov, $\neq$ Not Svetlanov), Gergev

Applying DR to the assertion protocol will yield the following protocol:

(11) Assertion with multiple responders

1. LatestMove = Assert(A,p)
2. A: push p? onto QUD; release turn
3. Resp_1: push p? onto QUD; take turn; $\langle$ Option 1: Discuss p?, Option 2: Accept p $\rangle$
4. Resp_2: push p? onto QUD; take turn; $\langle$ Option 1: Discuss p?, Option 2: Accept p $\rangle$
5. . . .
6. Resp_n: push p? onto QUD; take turn; $\langle$ Option 1: Discuss p?, Option 2: Accept p $\rangle$

One arguable problem with this protocol—equally applicable to the corresponding DRed grounding protocol—is that it licences long distance acceptance and is, thus, inconsistent with the MAG benchmark. On the other hand, it is potentially useful for interactions where there is explicitly more than one direct addressee.

A principle intermediate between AOV and DR is Add Side Participants (ASP):

(12) Given a dialogue protocol π, add roles $C_1,\ldots,C_n$, which effect the same contextual update as the interaction initiator.

Applying ASP to the dialogue assertion protocol yields the following protocol:

(13) Assertion for a conversation involving $\{A,B,C_1,\ldots,C_n\}$

1. LatestMove = Assert(A,p)
2. A: push p? onto QUD; release turn
3. C_i: push p? onto QUD;
4. B: push p? onto QUD; take turn; ⟨Option 1: Accept p, Option 2: Discuss p?⟩

(14)
1. LatestMove = Accept(B,p)
2. B: increment FACTS with p; pop p? from QUD;
3. C_i:increment FACTS with p; pop p? from QUD;
4. A: increment FACTS with p; pop p? from QUD;

This protocol satisfies the MAG benchmark in that acceptance is strictly local. This is because it enforces *communal acceptance*—acceptance by one CP can count as acceptance by all other addressees of an assertion. There is an obvious rational motivation for this, given the difficulty of a CP constantly monitoring an entire audience (when this consists of more than one addressee) for acceptance signals—it is well known that the effect of visual access on turn taking is highly significant (Dabbs and Ruback, 1987). It also enforces quick reaction to an assertion—anyone wishing to dissent from p must get their reaction in early i.e. immediately following the assertion since further discussion of $p?$ is not countenanced if acceptance takes place. The latter can happen of course as a consequence of a dissenter not being quick on their feet; on this protocol to accommodate such cases would require some type of backtracking.

Applying ASP to the dialogue querying protocol yields the following protocol:

(15) Querying for a conversation involving $\{$ A,B,C_1,...,$C_n$$\}$
1. LatestMove = Ask(A,q)
2. A: push q onto QUD; release turn
3. C_i: push q onto QUD;
4. B: push q onto QUD; take turn; make max-qud–specific utterance.

This improves on the DR generated protocol because it does allow responders to comment on previous responders—the context is modified as in the dialogue protocol. Nonetheless, as it stands, this protocol won't fully deal with examples such as (4)—the issue introduced by each successive participant takes precedence given that QUD is assumed to be a stack. This can be remedied by slightly modifying this latter assumption: we will assume that when a question q is pushed onto QUD it doesn't subsume *all* existing questions in QUD, but rather only those on which q does not depend:[7]

(16) q is QUD$^{mod(dependence)}$ maximal iff for any q_0 in QUD such that $\neg$Depend(q, q_1): $q \succ q_0$.

[7] The notion of dependence we assume here is one common in work on questions, e.g. (Ginzburg and Sag, 2000), intuitively corresponding to the notion of 'is a subquestion of'. q_1 depends on q_2 iff any proposition p such that p resolves q_2 also satisfies p is about q_1.

This is conceptually attractive because it reinforces that the order in QUD has an intuitive semantic basis. One effect this has is to ensure that any polar question $p?$ introduced into QUD, whether by an assertion or by a query, subsequent to a wh-question q on which $p?$ depends does not subsume q. Hence, q will remain accessible as an antecedent for NSUs, as long as no new unrelated topic has been introduced. Assuming this modification to QUD is implemented in the above ASP–generated protocols, both MLDSA and MAG benchmarks are fulfilled.

5 Conclusions and Further Work

In this paper we consider how to scale up dialogue protocols to multilogue, settings with multiple conversationalists. We have extracted two benchmarks, MLDSA and MAG, to evaluate scaled up protocols based on the long distance resolution possibilities of NSUs in dialogue and multilogue in the BNC. MLDSA, the requirement that multilogue protocols license long distance short answers, derives from the statistically significant increase in frequency of long distance short answers in multilogue as opposed to dialogue. MAG, the requirement that multilogue protocols enforce adjacency of acceptance and grounding interaction, derives from the overwhelming locality of acceptance/grounding interaction in multilogue, as in dialogue. In light of these benchmarks, we then consider three possible transformations to dialogue protocols formulated within an issue-based approach to dialogue management. Each transformation can be intuited as adding roles that correspond to distinct categories of an audience originally suggested by Goffman. The three transformations would appear to be complementary—it seems reasonable to assume that application of all three (in some formulation) will be needed for wide coverage of multilogue. MLDSA and MAG can be fulfilled within an approach that combines the Add Side Participants transformation on protocols with an independently motivated modification of the structure of QUD from a canonical stack to a stack where maximality is conditioned by issue dependence.

With respect to long distance short answers our account licences their occurrence in dialogue, as in multilogue. We offer a pragmatic account for their low frequency in dialogue, which indeed generalizes to explain a statistically significant correlation we observe between their increased incidence and increasing active participant size. We plan to carry out more detailed work, both corpus–based and experimental, in order to evaluate the status of MAG and, correspondingly to assess just how local acceptance and grounding interaction really are. We also intend to implement multilogue protocols in CLARIE so it can simulate multilogue. We will then evaluate its ability to process NSUs from the BNC.

Acknowledgements

We would like to thank three anonymous ACL reviewers for extremely useful comments, which in particular forced us to rethink some key issues. We would also like to thank Pat Healey, Shalom Lappin, Richard Power, and Matt Purver for discussion, and Zoran Macura and Yo Sato for help in assessing the NSU taxonomy. Earlier versions of this work were presented at colloquia at ITRI, Brighton, and at the Université Paris, 7. The research described here is funded by grant number RES-000-23-0065 from the Economic and Social Research Council of the United Kingdom.

References

Special issue on best practice in spoken language dialogue systems engineering. 2003. *Natural Language Engineering*.

Herbert Clark. 1996. *Using Language*. Cambridge University Press, Cambridge.

Robin Cooper. 2004. A type theoretic approach to information state update in issue based dialogue management. Invited paper, *Catalog'04, the 8th Workshop on the Semantics and Pragmatics of Dialogue*, Pompeu Fabra University, Barcelona.

James Dabbs and R. Barry Ruback. 1987 Dimensions of group process: amount and structure of vocal interaction. *Advances in Experimental Social Psychology* 20, pages 123–169.

Frank P.M. Dignum and Gerard A.W. Vreeswijk. 2003. Towards a testbed for multi-party dialogues. In *Proceedings of the first International Joint Conference on Autonomous Agents and Multi-agent Systems (AAMAS 2003)*.

Nicholas Fay, Simon Garrod, and Jean Carletta. 2000. Group discussion as interactive dialogue or serial monologue. *Psychological Science*, pages 481–486.

Raquel Fernández and Jonathan Ginzburg. 2002. Non-sentential utterances: A corpus study. *Traitement automatique des languages. Dialogue*, 43(2):13–42.

FIPA. 2003. The foundation for intelligent physical agents. interaction protocol specifications. http://www.fipa.org.

Roger Garside. 1987. The CLAWS word-tagging system, In Roger Garside et al. editors, *The computational analysis of English: a corpus-based approach*, Longman, Harlow, pages 30–41.

Jonathan Ginzburg and Robin Cooper. 2001. Resolving ellipsis in clarification. In *Proceedings of the 39th Meeting of the Association for Computational Linguistics*, Toulouse.

Jonathan Ginzburg and Ivan A. Sag. 2000. *Interrogative Investigations: the form, meaning and use of English Interrogatives*. Number 123 in CSLI Lecture Notes. CSLI Publications, Stanford: California.

Jonathan Ginzburg. (forthcoming). *Semantics and Interaction in Dialogue* CSLI Publications and University of Chicago Press.

Jonathan Ginzburg. 1996. Interrogatives: Questions, facts, and dialogue. In Shalom Lappin, editor, *Handbook of Contemporary Semantic Theory*. Blackwell, Oxford.

Erving Goffman 1981 *Forms of Talk*. University of Pennsylvania Press, Philadelphia.

Staffan Larsson. 2002. *Issue based Dialogue Management*. Ph.D. thesis, Gothenburg University.

Colin Matheson and Massimo Poesio and David Traum. 2000. Modelling Grounding and Discourse Obligations Using Update Rules. *Proceedings of NAACL 2000*, Seattle.

Stephen Pulman. 1997. Focus and higher order unification. *Linguistics and Philosophy*, 20.

Matthew Purver. 2004. *The Theory and Use of Clarification in Dialogue*. Ph.D. thesis, King's College, London.

David Traum and Jeff Rickel. 2002. Embodied agents for multi-party dialogue in immersive virtual world. In *Proceedings of the first International Joint Conference on Autonomous Agents and Multi-agent Systems (AAMAS 2002)*, pages 766–773.

David Traum. 2003. Semantics and pragmatics of questions and answers for dialogue agents. In H. Bunt, editor, *Proceedings of the 5th International Workshop on Computational Semantics*, pages 380–394, Tilburg. ITK, Tilburg University.